

Bitcoin Price Prediction Using Machine Learning

Abhinek Kumar Agarwal¹, Tushar Kumar², Shivam Raj³, Soham Morajkar⁴, Prof. Sagar Dhanake⁵

Student, Department of Computer Engineering^{1,2,3,4}

Assistant Professor, Department of Computer Engineering⁵

Dr. D. Y. Patil Institute of Technology, Pune, Maharashtra, India

Abstract: *With the increase of the digital cryptocurrency its shows abnormal profit in past few years. The price of bitcoin is fluctuating every second, which is why we keep track of it. want prediction model which give us a brief idea regarding the increase and decrease in bitcoin price. Here machine learning comes into scene by the help of machine learning and its LSTM and ARIMA models we get knowledge about the increase and decrease in the bitcoin price by their prediction models. In the first phase of our survey, we understand the daily trends of the bitcoin market through API as google, twitter, social media etc. In second phase we use the available data from API, and with the help of prediction models we predict the price of the bitcoin to the closest accuracy possible. Therefore, this paper includes the brief idea regarding Bitcoin price prediction system using Machine Learning Models.*

Keywords: Machine Learning, LSTM & RNN, ARIMA, Cryptocurrency

I. INTRODUCTION

Bitcoin is a decentralized digital currency, without a central bank or a single administrator, that can be sent directly from one user to another on the peer-to-peer bitcoin network. Network nodes use cryptography to verify transactions, which are then stored in a public distributed ledger called a blockchain. The cryptocurrency was invented in 2008 by an unknown person or group of people using the name Satoshi Nakamoto. When the currency's implementation was released as open-source software in 2009, it went into usage. In a white paper published on October 31, 2008, the term bitcoin was defined. It's made out of the terms "bit" and "coin". There is no universal capitalization convention for bitcoin; some sites use Bitcoin, capitalized, to refer to bitcoin, refers to the technology and network, with bitcoin (lowercase) serving as the unit of account. The Wall Street Journal, The Chronicle of Higher Education, and the Oxford English Dictionary advocate the use of lowercase bitcoin in all cases. The bitcoin is the bitcoin system's unit of account. Currency codes for representing bitcoin are BTC[a] and XBT. [b] [b] The Unicode character for it is. One bitcoin is divisible to eight decimal places. Units for smaller amounts of bitcoin are the milli bitcoin (MTBC), equal to 1/1000 bitcoin, and the Satoshi (sat), which is the smallest possible division, and named in homage to bitcoin's creator, representing 1/100000000 (one hundred million the) bitcoin. One MTBC is 100,000 Satoshi's. Bitcoin addresses are associated with bitcoins on the blockchain. Creating a bitcoin address requires nothing more than picking a random valid private key and computing the corresponding bitcoin address. This computation can be done in a half second. Computing the private key of a particular bitcoin address, on the other hand, is essentially impossible. Users can tell others or make public a bitcoin address without compromising its corresponding private key. Furthermore, the number of viable private keys is so large that it's quite unlikely that someone will compute a key pair that's already in use and has money. Due to the large number of valid private keys, brute force cannot be used to compromise a private key. The owner must know the appropriate private key and digitally sign the transaction in order to spend their bitcoins. [d] The network verifies the signature using the public key; the private key is never revealed. So, in today's world of growing market in cryptocurrencies especially Bitcoin we are going to predict the price of bitcoin with the help of machine learning models to make it easy for investors and big financial industries to buy bitcoin.

II. DESCRIPTION OF THE PROBLEM

Statement of Problem

Bitcoin is the most complex cryptocurrency which value change in every second. Investing in Bitcoin carries a higher risk and yields a lower return. So to make it easy for investors to invest in bitcoin and big financial industries to buy bitcoin by

the predicated price of bitcoin with the help of the Machine Learning models.

Existing System

There have been various commercial systems for BITCOIN price prediction via machine learning around the world. Two types of classifiers are used in these systems. They are GRU (Gated recurrent units) method and ARIMA (Autoregressive integrated moving average).

III. METHODOLOGY

To implement the model first we have to collect the data which means data gathering, after pre-processing and built model by the help of python library TensorFlow, scikit-learn. LSTM, ARIMA model and RNN models are used.

IV. LITERATURE REVIEW

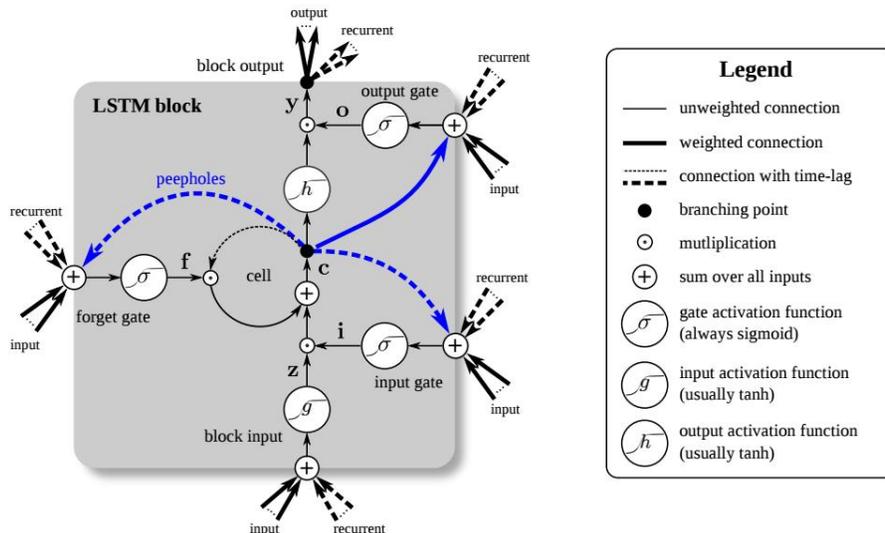
We used a journal publication on bitcoin price prediction using machine learning in this research. The articles chosen were among those published in recent years, while older articles were also included because some methods lost their popularity in recent times. The frequency of the dataset, the causation relationship, and the approach were all examined in the publications.

V. SYSTEM DESIGN AND FLOW

Long Short Term Memory (LSTM):

Long Short Term Memory networks LSTM are a special kind of RNN, capable of learning long-term dependencies. They were introduced by Hochester & Schmid Huber (1997), and were refined and popularized by many people in following work. They are currently frequently utilised and function exceptionally effectively on a wide range of situations. Unlike standard feed forward neural networks. LSTM has feedback connections. It can not only process individual data points, but also complete data sequences. LSTMs are explicitly designed to avoid the long-term dependency problem. They don't have to work hard to remember knowledge for lengthy periods of time; it's like second nature to them!

LSTM Architecture Diagram



LSTM Use Cases

Apple had been the first major tech company to integrate a smart assistant into its operating system. Siri was the company's adaptation of a standalone app it had purchased, along with the team that created it, in 2010. Over the past several years. On



some of the world's most demanding computer science challenges, deep learning has showed extraordinary accomplishment. Using techniques ranging from picture classification and captioning to translation and model visualisation, Recently Google announced improvements to "Google Voice transcription" using Long Short-term Memory Recurrent Neural Networks.

ARIMA (Autoregressive Integrated Moving Average)

An autoregressive integrated moving average (ARIMA), is a statistical analysis model that uses Time Series data for better understanding of the data set or to predict upcoming future trends. A statistical model is a autoregressive if it predicts future values which are based on past. For example, an ARIMA model will predict a stock's future prices based on its past performance or forecast a company's earnings form the data based on past. They are used widely in technical analysis to forecast future security prices. But sometimes they can prove inaccurate under certain market conditions, like a financial crises or periods of rapid technological change.

Arima Model Specification

ARIMA (p,d,q) tells us the number of lags of the dependent variable(p), how many times the variable is differenced to become stationary (d) and the number lags of the error term(q) such that:

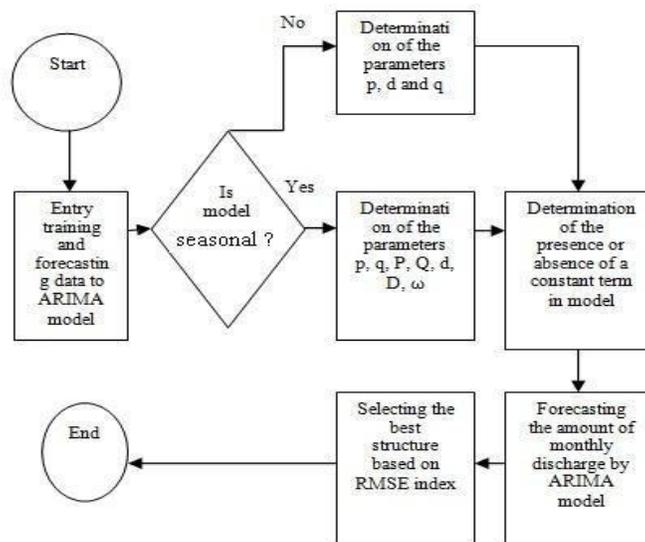
1. ARIMA(1,1,2) indicates the model has: one lag of the dependent variable (1). the variable being used is of first-difference stationary (1) and two lags of the error term (2)
2. ARIMA(1,2,1) indicates the model has : one lag of dependent variable (1), the variable being used is of level stationary (0) and one lag of the error term(1).
3. ARIMA(1,0,1) = ARMA(1,1) if the series is stationary in level.

ARIMA Parameters

Each component of ARIMA is represented as a parameter using a standard syntax. For models like ARIMA, a standard notation would be ARIMA with p, d, and q, where integer values substitute for the parameters to indicate the type of ARIMA model used. The parameters can be defined as:

- p: refers to the order of the autoregressive term.
- d: is the order of differencing in order to make the time series stationary.
- q: is the order of the moving average term.

ARIMA Architecture Diagram



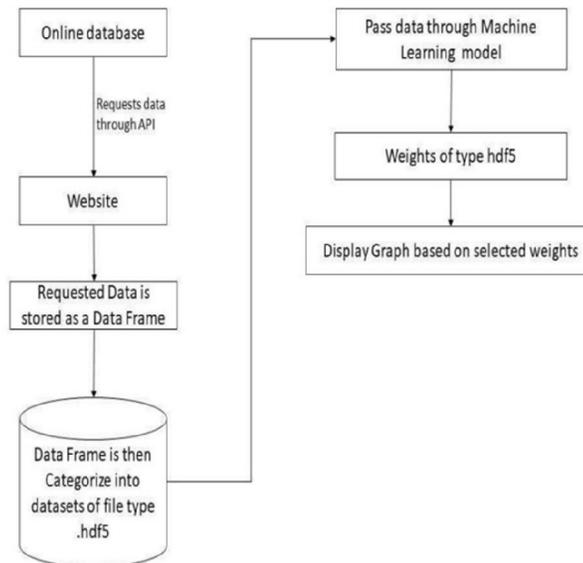


Figure: Flow Diagram

VI. PROJECT IMPLEMENTATION

Data Pre-processing

The primary knowledge collected from the web sources remains within the raw kind of statements, digits and qualitative terms. The raw data contains error, omissions and inconsistencies. It requires corrections after careful scrutinizing the completed questionnaires. The following steps square measure concerned within the process of primary knowledge. A huge volume of information collected through field survey must be sorted for similar details of individual responses. Data Pre-processing could be a technique that's accustomed convert the {raw knowledge| data| information} into a clean data set. In alternative words, whenever the info is gathered from totally different sources it's collected in raw format that isn't possible for the analysis. Therefore, bound steps square measure dead to convert {the knowledge| the info| the information} into a little clean data set. This technique is performed before the execution of reiterative Analysis. The set of steps is understood as knowledge pre-processing. The process comprises:

- Data Gathering
- Data Cleaning
- Data Normalization

Data Gathering

First, the Bitcoin price history, which is extracted from Coin market cap through its open API . Secondly, data from Blockchain is gathered, in particular we choose the average block size, the number of user addresses, number of transactions, and the miners revenue.

Data Cleaning

From exchange data we consider relevant only the Volume, Close, Open, High prices and Market capitalization. For all data sets if NaN values are found to be existent, they are replaced with the mean of the respective attribute. Then, along the time dimension, all datasets are blended into one.



Data Normalization

We need to normalize inputs in neural networks and other data mining models, otherwise, the network will perform poorly. Normalization is used to ensure that each RNN model input has the same range of values. This ensures that weight and partiality convergence remain steady. Normalization here uses the Minmax Scalar Package, after normalization, data is plotted using mat plot libraries and the trend is used to examine the price and volume changes of bitcoin over the last two years.

Model Implementation



Figure: 2017 – 2022 Bitcoin close price range

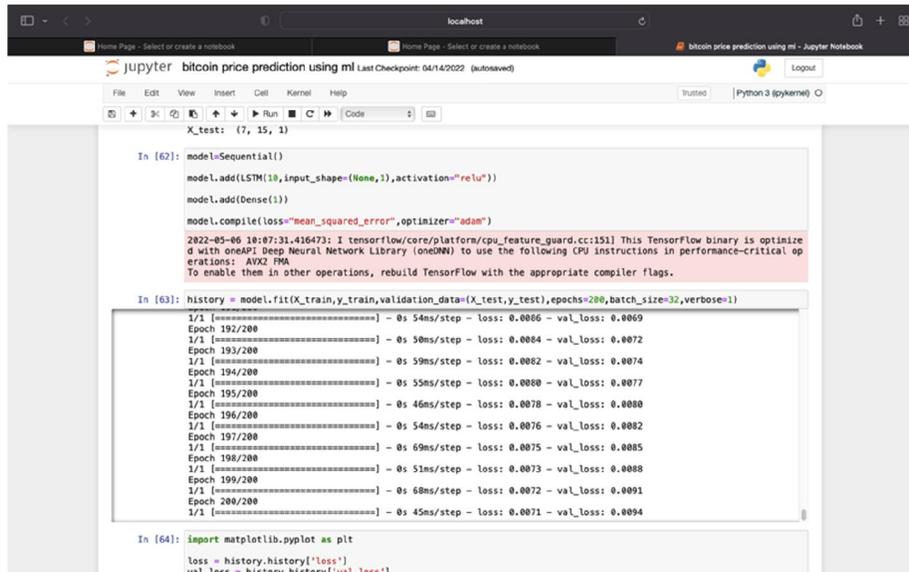


Fig. Epoch of the Model

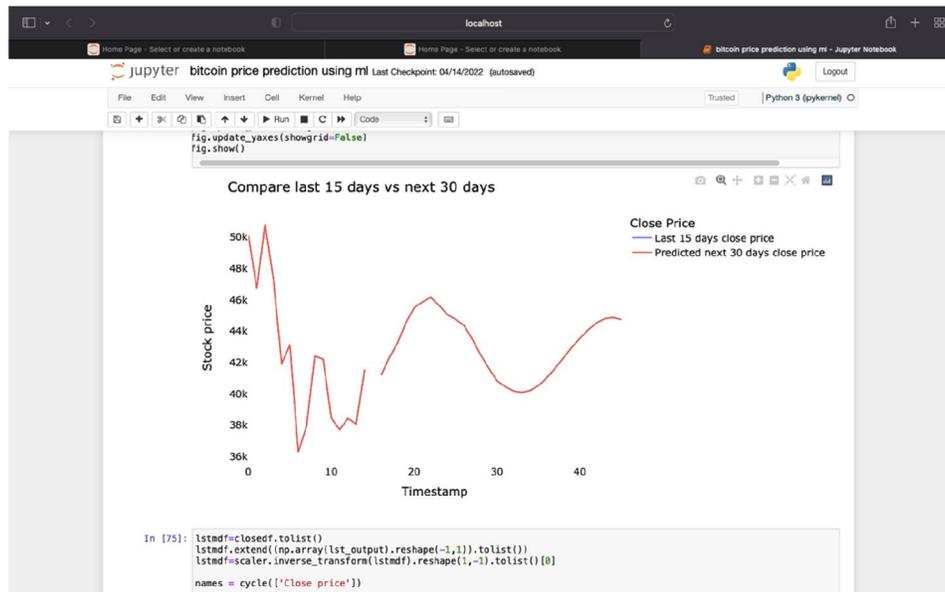


Fig. Compare last 15 days vs next 30 days Price

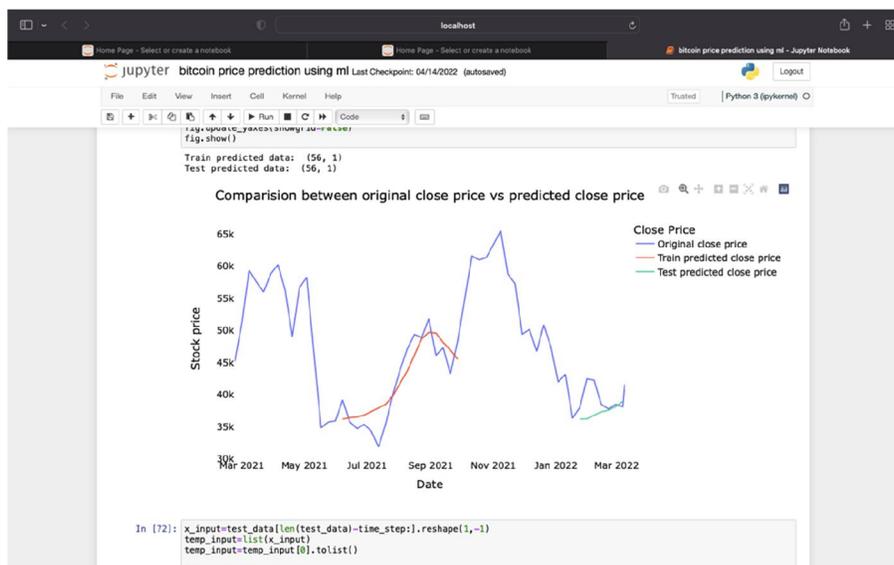


Fig. comparison between org. close price vs predicted close price

Splitting of data:

We divide the data into training and test data in statistics and machine learning. The model is intended to fit the training data to make predictions. When we do this, there are chances of two things happening, one over fitting the model and the other under fitting the model. Over fitting means that the model is too well trained, and the predictions are too close and that it does not fit closely with the model when the model is under fitting. To separate the data, the Scikit Library is used. The data for training is split into two parts: 70% for training and 30% for testing.

The Root Mean Squared Error:

Time series usually focus on predicting real values, which are called regression problems. Hence, the performance measurements in this tutorial focus on real-value prediction evaluation methods. The most commonly used mean squared error (MSE), root mean squared error (RMSE).

VII. DISCUSSION AND RESULTS

In order to decide proper strategy of analysis for studying the connection amongst Bitcoin's The existing literature has been thoroughly examined in terms of price and other key metrics, as well as sentiment analysis. The majority of articles [19], [20], [21] discuss the existing relationship between the amount of tweets and market evolution. Bolen et al., in general, demonstrated that tweets are expecting the market trend 3-4 days in advance, with a very good chance of success from our experiments. The result shows that machine learning models (LSTM) take much longer to compile because of their complex calculations than traditional models (ARIMA). The LSTM model compilation time is 61 milliseconds and 4 milliseconds for the ARIMA model.

VIII. ADVANTAGES & DISADVANTAGES

Advantages:-

Protection from inflation Inflation has caused many currencies to get their value declined with time. Almost every cryptocurrency, at the time of its launch, is released with a fixed amount. The source code specifies the amount of any coin; like, there are only 21 million Bitcoins released in the world. So, as the demand increases, its value will increase which will keep up with the market and, in the long run, prevent inflation.

Self-governed and managed – Governance and maintenance of any currency is a major factor for its development. The cryptocurrency transactions are stored by developers/miners on their hardware, and they get the transaction fee as a reward for doing so. Since the miners are getting paid for it, they keep transaction records accurate and up-to-date, keeping the integrity of the cryptocurrency and the records decentralized.

- Secure and private – Privacy and security have always been a major concern for cryptocurrencies. The blockchain ledger is based on different mathematical puzzles, which are hard to decode. This makes a cryptocurrency more secure than ordinary electronic transactions. Cryptocurrencies, for better security and privacy, use pseudonyms that are unconnected to any user, account or stored data that could be linked to a profile.
- Currency exchanges can be done easily – Cryptocurrency can be bought using many currencies like the US dollar, European euro, British pound, Indian rupee or Japanese yen. With the help of different cryptocurrency wallets and exchanges, one currency can be converted into the other by trading in cryptocurrency, across different wallets, and with minimal transaction fees.
- Decentralized – A major pro of cryptocurrency is that they are mainly decentralized. A lot of cryptocurrencies are controlled by the developers using it and the people who have a significant amount of the coin, or by an organization to develop it before it is released into the market. The decentralization helps keep the currency monopoly free and in check so that no one organization can determine the flow and the value of the coin, which, in turn, will keep it stable and secure, unlike fiat currencies which are controlled by the government.
- Cost-effective mode of transaction – One of the major uses of cryptocurrencies is to send money across borders. With the help of cryptocurrency, the transaction fees paid by a user is reduced to a negligible or zero amount. It does so by eliminating the need for third parties, like VISA or PayPal, to verify a transaction. This removes the need to pay any extra transaction fees.

Disadvantages:-

- Can be used for illegal transactions – Since the privacy and security of cryptocurrency transactions are high, it's hard for the government to track down any user by their wallet address or keep tabs on their data. Bitcoin has been used as a mode of exchanging money in a lot of illegal deals in the past, such as buying drugs on the dark web. Cryptocurrencies are also used by some to convert their illicitly obtained money through a clean intermediary, to hide its source.

- Data losses can cause financial losses – The developers wanted to create virtually untraceable source code, strong hacking defences, and impenetrable authentication protocols. This would make it safer to put money in cryptocurrencies than physical cash or bank vaults. But if any user loses the private key to their wallet, there's no getting it back. The wallet will remain locked away along with the number of coins inside it. This will result in the financial loss of the user.
- Decentralized but still operated by some organization – The cryptocurrencies are known for its feature of being decentralized. But, the flow and amount of some currencies in the market are still controlled by their creators and some organizations. These holders can manipulate the coin for large swings in its price. Even hugely traded coins are susceptible to these manipulations like Bitcoin, whose value doubled several times in 2017.
- Some coins not available in other fiat currencies – Some cryptocurrencies can only be traded in one or a few fiat currencies. This forces the user to convert these currencies into one of the major currencies, like Bitcoin or Ethereum first and then through other exchanges, to their desired currency. This applies to only a few cryptocurrencies. By doing this, the extra transaction fees are added in the process, costing unnecessary money.
- Adverse Effects of mining on the environment – Mining cryptocurrencies require a lot of computational power and electricity input, making it highly energy-intensive. The biggest culprit in this is Bitcoin. Mining Bitcoin requires advanced computers and a lot of energy. It cannot be done on ordinary computers. Major Bitcoin miners are in countries like China that use coal to produce electricity. This has increased China's carbon footprint tremendously.
- Susceptible to hacks – Although cryptocurrencies are very secure, exchanges are not that secure. Most exchanges store the wallet data of users to operate their user ID properly. This data can be stolen by hackers, giving them access to a lot of accounts.

After getting access, these hackers can easily transfer funds from those accounts. Some exchanges, like Bitfinex or Mt Gox, have been hacked in the past years and Bitcoin has been stolen in thousands and millions of US dollars. Most exchanges are highly secure nowadays, but there is always a potential for another hack.

IX. CONCLUSION & FUTURE WORK

This study focuses on the Bitcoin closing price and sentiments of the current market for the development of the predictive model. It also calculates market sentiments in order to better anticipate the price. The prediction is limited to previous data. The ability to predict data streaming would improve the model's performance and predictability. So with the help of the bitcoin price prediction system models we get closely predicted price of the bitcoin.

Bitcoin can be seen as the most commonly used digital currency today, attracting investor's attention (Blau 2018). Bitcoin's value has skyrocketed in recent years, resulting in increased demand. Its recognition by reputable companies such as Amazon, Microsoft, Overstock, DISH Network, Intuit, and even PayPal. In countries such as Japan, Netherlands, Canada and the United States, you can pay Bitcoin at restaurants, malls and other large and small businesses. On the other hand, the prediction of Bitcoin price is not an easy task since it is a new and unstable market. Because the grey system theory can create predictions based on a tiny set of data, it's a good idea to use it. We utilized this strategy to forecast Bitcoin price for the next day based on incomplete data and incomplete information.

ACKNOWLEDGEMENTS

The completion of our project brings with it a sense of satisfaction, but it is never complete without those people who made it possible and whose constant support has crowned our efforts with success. One cannot even imagine our completion of the project without guidance and neither can we succeed without acknowledging it. It is a great pleasure that we acknowledge the enormous assistance and excellent co-operation to us by the respected personalities.

REFERENCES

- [1]. Nakamoto S. Bitcoin is an electronic monetary system that works on a peer-to-peer basis. <https://bitcoin.org/bitcoin.pdf>; Work Paper; 2008.
- [2]. R. Beck, M. Avital, M. Rossi, and J.B. Thatcher. Blockchain technology in business and information systems research. *Bus Inf System Eng.* 2017;59:381e384. <https://doi.org/10.1007/s12599-017-0505-1>.

- [3]. S. Gu, B. Kelly, and D. Xiu. Empirical asset pricing via machine learning. *Rev Finance Stud.* 2020;33:2223e2273. <https://doi.org/10.1093/rfs/hhaa009>.
- [4]. G. Feng, S. Giglio, and D. Xiu. Taming the factor zoo: a test of new factors. *J Finance.* 2020;75:1327e1370. <https://doi.org/10.1111/jofi.12883>.
- [5]. Jaquart P, Dann D, Martin C. Machine learning for bitcoin pricing - a structured literature review. In: *WI 2020 Proceedings*, GITO Verlag. 2020:174e188. https://doi.org/10.30844/wi_2020_b4-jaquart.
- [6]. Coinmarketcap; 2020. URL: <https://coinmarketcap.com/>. Accessed September 15, 2020.
- [7]. Fama EF. A review of theory and empirical studies on efficient capital markets. *J Finance*, vol. 25, no. 4, <https://doi.org/10.1111/j.1540-6261.1970.tb00518.x>.
- [8]. M. Dixon, D. Klabjan and J. H. Bang, "Classification-based financial markets prediction using deep neural networks," *ArXiv*, 2017.
- [9]. S. McNally, J. Roche and S. Caton, "Predicting the price of Bitcoin using machine learning," in *26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)*, 2018.
- [10]. M. Daniela and A. BUTOI, "Data mining for price prediction in the Romanian stock market using neural networks," *Informatica Economica*, vol. 17, no. 3, 2013.
- [11]. D. Shah and K "Bayesian regression and Bitcoin," in *52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2015. Zhang, "Bayesian regression and Bitcoin," in *52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2015.
- [12]. "An Empirical Study on Modeling and Prediction of Bitcoin Prices Using Bayesian Neural Networks Based on Blockchain Information," according to the paper by H. Jang and J. Lee, in *IEEE Early Access Articles*, 2017.
- [13]. F. A. d. Oliveira, L. E. Zarate, M. d. A. Reis and C.N. Nobre, "The use of artificial neural networks in the analysis and prediction of stock prices," in *IEEE International Conference on Systems, Man, and Cybernetics*, 2011.
- [14]. Available: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>.
- [15]. "High throughput quaternary signed digital adder design for portable electronic applications," *International Journal of Pure and Applied Mathematics*, vol. 116, no. 11, 2017, pp. 61-69. S. Saravanakumar and V.D. Kumar, "High throughput quaternary signed digital adder design for portable electronic applications," *International Journal of Pure and Applied Mathematics*, vol. 116, no. 11, 2017, pp. 61-69.